Kaijing MA

Tel: +86 18964475073 | **Email:** 2151400@tongji.edu.cn
Website: https://mkj3085003.github.io/

Research Interests: AI & NLP, LLM, Reasoning, Evaluation, Mech. Interpretability

EDUCATION

Tongji University **Shanghai, China**
Bachelor of Computer Science and Technology *Sep 2021 - Jun 2025*

- **GPA:** 92.19 / 100
- **Honors:** National Encouragement Scholarship for Undergraduate Students (2021-2024); First-class Scholarship for Outstanding Undergraduate Students (2021-2022); 2023 Tongji University Inspirational Star Award Nomination

PUBLICATIONS

- **Scaling Latent Reasoning via Looped Language Models.** *Mechanistic Interpretability*
- **Criticlean: Critic-guided reinforcement learning for mathematical formalization.** ***Co-author***
- **Seed-prover: Deep and broad reasoning for automated theorem proving**. *Data Support*
- **KORGym: A Dynamic Game Platform for LLM Reasoning Evaluation.** ***(NeurIPS 2025 Spotlight)*** *Game Support*
- **SuperGPQA: Scaling LLM Evaluation across 285 Graduate Disciplines.*(NeurIPS 2025)*** ***Leading Author***
- **KOR-Bench: Benchmarking Language Models on Knowledge-Orthogonal Reasoning Tasks.** ***(ICLR 2025)*** ***First Author***
- **KARPA: A Training-free Method of Adapting Knowledge Graph as References for Large Language Model's Reasoning Path Aggregation.** ***(ACL Findings 2025)*** *Second author*
- **CodeEditorBench: Evaluating Code Editing Capability of Large Language Models.** ***(ICLR 2025 DL4C)*** ***Co-author***
- **SciMMIR: Benchmarking Scientific Multi-modal Information Retrieval.** ***(ACL Findings 2024)*** *Data Support*
- **MAP-Neo: Highly Capable and Transparent Bilingual Large Language Model Series.** *Data Pipeline*

ACADEMIC EXPERIENCE

OpSynth-MI (Operator Synthesis for Mechanistic Interpretability) *Oct 2024 - Present*

- Built a controllable operator-expression generation system, defining novel recursive and compositional mathematical operators and automatically producing large-scale expression corpora for LLM pretraining.
- Structured the data with explicit mathematical dependencies, enabling systematic tracing and retrieval of the model's reasoning chains.
- Applied mechanistic interpretability methods to analyze how models learn and generalize.

KOR-Bench: Benchmarking Language Models on Knowledge-Orthogonal Reasoning Tasks *Jun 2024 - Jan 2025*
Github: https://github.com/KOR-Bench/KOR-Bench | Website: https://kor-bench.github.io/

- Created KOR-Bench, an innovative benchmark accepted to ICLR 2025 that evaluates LLMs' reasoning in novel domains orthogonal to pretraining knowledge.
- Designed 5 specialized reasoning categories (Operation, Logic, Cipher, Puzzle, Counterfactual) with unique rule systems to systematically assess models' adaptability to unfamiliar problem structures.
- Developed the complete benchmark framework, including data construction, evaluation methodology, and co-authored the research paper documenting its design and significance.

CodeEditorBench: Evaluating Code Editing Capability of Large Language Models *Dec 2023 - Jun 2024*
Github: https://github.com/CodeEditorBench/CodeEditorBench | Website: https://codeeditorbench.github.io/

- Developed CodeEditorBench, a comprehensive benchmark evaluating LLMs on diverse real-world code editing tasks across multiple programming languages, assessing 17 models to reveal performance differences.
- Designed and implemented an automated Online Judge (OJ) system for evaluation of code editing accuracy.
- Published open-source prompts, datasets, and documentation to foster community adoption and further research in LLM-based code editing.

PROFESSIONAL EXPERIENCE

ByteDance

Intern, advised by Dr. Wenhao Huang and Dr. Ge Zhang *May 2025 - Oct 2025*

- Led OpSynth-MI, a controllable operator-expression generation framework with novel operators, enabling scalable pretraining and mechanistic interpretability studies of LLM reasoning and generalization.
- Co-authored CriticLean and built FineLeanCorpus, a semantically validated dataset of 285K math problems with Lean code, supporting a generate – critic – revise RL loop for formalization.
- In Seed-Prover, developed the Lean proof dataset enabling lemma-centric reasoning and integration with the Seed-Geometry engine for neural – symbolic geometric proofs.

MIT CSAIL

Intern, advised by Prof. Wojciech Matusik *June 2025 - Present*

- Developed MusicDSL, a domain-specific language for expressing musical structure and linking DAWs.
- Built middleware enabling bidirectional communication between MusicDSL, DAWs, and AI models.
- Fine-tuned generative audio and symbolic-music models on MusicDSL-aligned data to improve controllability and structural consistency.

Multimodal Art Projection (MAP)

Intern *Oct 2023 - Present*

- Conducted advanced research on LLMs, focusing on their reasoning capabilities, task adaptability, and code editing proficiency
- Contributed to the development of CodeEditorBench (a benchmark for real-world code editing) and KOR-Bench (a framework for testing orthogonal reasoning tasks), enhancing model evaluation methodologies
- Authored detailed research reports covering experimental design, data analysis, results interpretation, and key conclusions to support team projects and future publications

Shanghai Research Institute for Intelligent Autonomous Systems

Research Assistant *Jun 2022 - Sep 2023*

- Took charge of writing the patent project "Cloud-edge Collaborative-Aware Multi-Terminal Cross-Space-Time Pedestrian Re-Identification Method and System" under the guidance of supervisor
- Proposed a two-level pedestrian re-identification clustering coding method for cloud edge co-perception, and implemented a scalable and evolutionary continuous learning pedestrian re-identification algorithm

SPAR Project — Mentee *Feb 2024 – Jun 2024*

- Implemented a secure steganography system integrating iMEC and GPT-2.

AI Safety Hungary Technical Course — Trainee *Feb 2024 – Apr 2024*

- Studied AI alignment and safety through technical readings and discussions.

COMPETITION AWARDS

- Excellence Award, 2023 CCF Software Conference Robotics Large Model and Embodied Intelligence Competition
- First Prize, Professional Track 1, 2023 AI for Brain Science Collegiate Challenge
- First Prize, Creative Group, 2023 Shanghai Female Student Innovation and Entrepreneurship Competition
- Ranked 4th, 2023 VEX Robotics World Championships VEX U Design Division
- Design Award, 2023 China University Students Intelligent Robot Creativity Competition

LANGUAGES & SKILLS

Languages: Chinese (Native); English (Fluent)

Programming Languages: C/C++, Python,Verilog, JavaScript, HTML/CSS, Assembly Language

Robotics & Simulation: Hardware & Mechanical Skills: Mechanical assembly, SolidWorks, 3D printing; Software & Simulation: ROS, Sensors, Basic control

Machine Learning & Deep Learning: PyTorch, TensorFlow, Hugging Face Transformers

Large Language Models: Inference: vLLM, Pretraining: Megatron-LM, Fine-tuning (SFT): LLaMA-Factory, Reinforcement Learning (RL): VERL, Mechanistic Interpretability: NNSight, Distributed Computing: Ray